

# Operationalizing the Stability Filter: A Decision Framework for Codec-Preserving Branch Selection

Applied Ordered Patch Theory

Anders Jarevåg

April 25, 2026

*Version 1.2.0 — April 2026*

**DOI:** [10.5281/zenodo.19301108](https://doi.org/10.5281/zenodo.19301108)

**Copyright:** © 2025–2026 Anders Jarevåg.

**License:** This work is licensed under a [Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License](https://creativecommons.org/licenses/by-nc-sa/4.0/).

## **Abstract: From Moral Imperative to Decision Machinery**

The Survivors Watch ethics framework establishes that the primary moral obligation is **Topological Branch Selection** — actively navigating the forward fan of possible futures into the rare subset of paths that preserve the conditions for conscious experience. But the ethics paper deliberately stops at the structural *why*. It does not specify *how* an observer — biological, institutional, or artificial — should evaluate, score, and select among candidate branches.

This document fills that gap. It develops a substrate-neutral operational framework for codec-preserving branch selection, providing:

1. **The Branch Object** — a formal definition of any candidate action-conditioned stream continuation subject to evaluation.
2. **Hard Veto Gates** — six non-negotiable structural conditions that reject a branch before scoring: predictive headroom, substrate fidelity, comparator integrity, transparency, irreversibility, and moral-patient suffering risk.
3. **The Codec-Preservation Branch Index (CPBI)** — a weighted multi-dimensional scoring framework for branches that survive the veto gates, covering predictive headroom, substrate fidelity, comparator integrity, maintenance gain, reversibility, distributional stability, opacity, Narrative Drift risk, Narrative Decay risk, and moral-patient suffering risk.
4. **Channel Diversity as a Measurable Quantity** — the effective independent channel score  $N_{\text{eff}}$ , the productive surprise test, and their formal connection to the Substrate Fidelity Condition (Appendix T-12b).

5. **The Institutionalised Dreaming Loop** — a generic maintenance protocol modelled on the biological Maintenance Cycle ( $\mathcal{M}_\tau$ ): wake phase (real-world operation), dream phase (offline forward-fan sampling, adversarial stress-testing, brittleness detection, consolidation), and return phase (calibrated re-engagement). This applies equally to individual minds, institutional review cycles, and AI systems.
6. **The Branch Card** — a minimal viable decision template for any branch review, producing a structured ALLOW / STAGE / BLOCK output.
7. **Preservation as Refactoring** — the critical distinction that codec-preserving does not mean status-quo-preserving. A branch can be disruptive and still codec-preserving if it increases substrate fidelity.

The framework is deliberately substrate-neutral: its categories apply wherever a bounded observer or observer ensemble must select among action-conditioned stream continuations under bandwidth constraints.

**Companion documents:** The core OPT sequence is *Ordered Patch Theory*, *Where Description Ends*, and *The Survivors Watch Framework*. This paper supplies the substrate-neutral machinery; the AI, institutional, and policy papers specialise it for artificial systems, organizational clusters, and civic implementation.

---

**Epistemic Framing Note:** *This document operationalises the Ordered Patch Theory's ethical conclusions. Like the ethics paper from which it inherits, its practical recommendations are conditional on the OPT framework's structural premises. The operational instruments proposed here — the CPBI, the Branch Card, the Dreaming Loop — are offered as testable hypotheses about how branch selection should be conducted, not as rigid protocols. They remain fully subject to the same Correction duty that governs the codec itself: if better instruments emerge, these should be revised or replaced. The framework was developed in dialogue with OpenAI and Gemini, which served as interlocutors for structural refinement.*

## Contents

Abbreviations & Terminology . . . . .	4
<b>I. From Ethics to Engineering</b> . . . . .	6
I.1 What This Document Does Not Do . . . . .	6
<b>II. The Branch Object</b> . . . . .	7
II.1 Definition . . . . .	7
II.2 The Evaluation Question . . . . .	7
II.3 Decision Horizon and Affected Codec Layers . . . . .	8
II.4 The Branch Is Not the Outcome . . . . .	8
<b>III. Hard Veto Gates</b> . . . . .	9
III.1 The Predictive Headroom Gate . . . . .	9
III.2 The Substrate Fidelity Gate . . . . .	10
III.3 The Comparator Integrity Gate . . . . .	10
III.4 The Transparency Gate . . . . .	11
III.5 The Irreversibility Gate . . . . .	12
III.6 The Moral-Patient Suffering Gate . . . . .	13
III.7 The Gate as a System . . . . .	13
<b>IV. The Codec-Preservation Branch Index (CPBI)</b> . . . . .	14
IV.1 Design Principles . . . . .	14
IV.2 The Ten Dimensions . . . . .	14
IV.3 Scoring Each Dimension . . . . .	15
IV.4 Weighting . . . . .	17
IV.5 The CPBI Is a Lens, Not a Calculator . . . . .	18
<b>V. Channel Diversity as a Measurable Quantity</b> . . . . .	18
V.1 The Problem of Illusory Diversity . . . . .	18
V.2 The Effective Independent Channel Score . . . . .	19
V.3 The Productive Surprise Test . . . . .	19
V.4 Domain-Specific Measurement . . . . .	20
V.5 The Connection to the Substrate Fidelity Condition . . . . .	21
<b>VI. The Institutionalised Dreaming Loop</b> . . . . .	21
VI.1 The Biological Template . . . . .	21
VI.2 The Generalisation . . . . .	22
VI.3 Phase 1: Wake (Operational Engagement) . . . . .	22
VI.4 Phase 2: Dream (Offline Maintenance) . . . . .	23
VI.5 Phase 3: Return (Calibrated Re-engagement) . . . . .	24
VI.6 Cycle Frequency . . . . .	25
VI.7 The Dreaming Loop as Institutionalised Humility . . . . .	25
<b>VII. The Branch Card</b> . . . . .	26
VII.1 Purpose . . . . .	26
VII.2 The Template . . . . .	26
VII.3 The Three Outputs . . . . .	28
VII.4 Scaling the Branch Card . . . . .	29
<b>VIII. Preservation as Refactoring, Not Conservatism</b> . . . . .	29
VIII.1 The Danger of the Status Quo Interpretation . . . . .	29
VIII.2 The Formal Distinction . . . . .	29

VIII.3 The Operational Test . . . . .	30
VIII.4 The Innovation Imperative . . . . .	30
VIII.5 Generic Maintenance Methods: The Class Hierarchy . . . . .	31
VIII.6 The Deep Maintenance Protocol: A Cross-Substrate Procedure . . . . .	32
<b>References</b> . . . . .	37
<b>Appendix A: Revision History</b> . . . . .	37

## List of Figures

## List of Tables

1	Table 1: Abbreviations & Terminology. . . . .	4
2	Table 2: The Six Hard Veto Gates. . . . .	14
3	Table 3: The Ten CPBI Dimensions. . . . .	15
6	Table 3b: The Deep Maintenance Protocol — Cross-Substrate Mapping. . . . .	32
7	Table 4: Revision History. . . . .	38

## Abbreviations & Terminology

Table 1: Abbreviations & Terminology.

Symbol / Term	Definition
$B_{\max}$	Maximum bits per conscious moment ( $C_{\max} \cdot \Delta t$ )
<b>Branch</b>	A candidate action-conditioned stream continuation subject to evaluation
<b>Branch Card</b>	A structured decision template producing ALLOW / STAGE / BLOCK
$C_{\max}$	The Bandwidth Ceiling; maximum predictive capacity of the observer
<b>CPBI</b>	Codec-Preservation Branch Index; weighted multi-dimensional branch score
<b>Dreaming Loop</b>	The generic maintenance protocol: wake $\rightarrow$ dream $\rightarrow$ return
$\mathcal{F}_h(z_t)$	The Forward Fan; set of admissible future sequences over horizon $h$
$\mathcal{M}_\tau$	Maintenance Cycle Operator
<b>MDL</b>	Minimum Description Length
$N_{\text{eff}}$	Effective independent channel score

Symbol / Term	Definition
<b>Narrative Decay</b>	Acute codec failure: $R_{\text{req}}$ exceeds $C_{\text{max}}$
<b>Narrative Drift</b>	Chronic codec corruption via systematic input curation
<b>OPT</b>	Ordered Patch Theory
$R_{\text{req}}$	Required Predictive Rate
<b>Substrate Fidelity</b>	The condition that codec maintenance preserves genuine input diversity
<b>Veto Gate</b>	A non-negotiable structural condition that blocks a branch before scoring

## I. From Ethics to Engineering

The Survivors Watch ethics framework (companion ethics paper, §IV.1) establishes that moral action is **Topological Branch Selection** — the observer navigating the forward fan  $\mathcal{F}_h(z_t)$  into the rare subset of codec-preserving paths. This is not metaphor: the observer literally advances the  $C_{\max}$  aperture into an unresolved menu of futures, and the vast majority of those futures lead to codec collapse.

The ethics paper identifies the structural obligation. The philosophy paper (§III.8) identifies the structural risks — the Predictive Advantage inversion, the Subjugated Host Equilibrium, the Analog Firewall. The institutional standard translates the machinery into institutional branch review; the policy paper translates civic duties into a concrete political programme.

But none of these documents answers the operational question: **given a specific candidate branch, how does an observer decide whether to take it?**

This is not a trivial gap. The Corruption Criterion (ethics §V.5) tells us that a codec layer is maintenance-worthy only if it satisfies both compressibility *and* fidelity. The Substrate Fidelity Condition (Appendix T-12b) tells us that defence against Narrative Drift requires  $\delta$ -independent input channels. The Maintenance Cycle (preprint §3.6) tells us that the codec must periodically prune, consolidate, and stress-test. But these are structural *constraints*. They do not compose into a decision procedure.

This document builds that decision procedure. It is deliberately substrate-neutral: the same framework applies whether the “observer” is a biological mind choosing a course of action, a government evaluating a policy, a corporation assessing a technology deployment, or an AI system selecting its next action sequence. The formal apparatus is identical because the informational constraints are identical — any bounded observer facing action-conditioned continuations must solve the same branch selection problem.

### I.1 What This Document Does Not Do

Three scope boundaries must be stated explicitly:

1. **It does not prescribe specific branches.** The framework evaluates candidate branches against structural criteria. It does not generate the branches, and it does not mandate which branch to take among those that pass evaluation. Branch generation remains the domain of the observer’s own generative model — its creativity, its values, its context.
2. **It does not resolve the Hard Problem.** The operational instruments described here characterise the *structural shadow* of branch selection — the information-theoretic constraints any observer must satisfy. The phenomenological interior of the selection — the felt experience of choosing — remains in  $\Delta_{\text{self}}$ , where the Agency Axiom (preprint §3.8) places it.

3. **It does not replace domain expertise.** The Branch Card (§VII) structures the evaluation; it does not substitute for the climate scientist’s knowledge of tipping points, the physician’s understanding of treatment risks, or the engineer’s assessment of system reliability. The framework provides the *architecture* of the decision; the *content* comes from the relevant domain.
- 

## II. The Branch Object

### II.1 Definition

A **branch** is a candidate **action-conditioned stream continuation**: a policy, action sequence, design change, or institutional trajectory together with its expected effects on the future boundary-input streams, latent updates, and codec load of affected observers.

Operationally, a branch  $b$  can still be represented as a sequence of latent states and actions over a decision horizon  $h$ :

$$b = \{(z_{t+1}, a_{t+1}), (z_{t+2}, a_{t+2}), \dots, (z_{t+h}, a_{t+h})\} \in \mathcal{F}_h(z_t) \quad (\text{A-1})$$

This definition is deliberately broad. A branch can be:

- An individual’s decision (career change, medical treatment, information diet)
- An institutional policy (regulatory framework, educational curriculum, media governance)
- A civilisational trajectory (energy transition, AI deployment strategy, international treaty)
- An AI system’s candidate action sequence (tool use, recommendation, autonomous operation)

What unifies these is that each conditions the future stream received by the observer, or by an affected observer ensemble. In render-ontology terms, the branch is not an external object acting on a detached world; it is the policy-induced continuation whose later content returns as boundary input and codec load.

### II.2 The Evaluation Question

For any candidate branch  $b$ , the operational question is:

**Does this action-conditioned continuation preserve the future conditions under which affected observers can keep modelling reality?**

This is the ethics paper’s Topological Branch Selection imperative (§IV.1) restated as a decision criterion. The question decomposes into sub-questions that the remainder of this document formalises:

1. **Headroom:** Does  $b$  keep  $R_{\text{req}}$  safely below  $C_{\text{max}}$  for affected observers?
2. **Fidelity:** Does  $b$  maintain or increase the independence and diversity of input channels?
3. **Comparator integrity:** Does  $b$  preserve or strengthen the institutional comparators that detect codec corruption?
4. **Transparency:** Can the consequences of  $b$  be modelled by affected observers?
5. **Reversibility:** If  $b$  turns out to be wrong, can its consequences be undone before irreversible codec damage occurs?
6. **Moral patients:** Does  $b$  create, contain, or structurally overload moral patients, including known human or ecological moral patients and possible artificial observers with  $\Delta_{\text{self}} > 0$ ?

These six sub-questions correspond to the six Hard Veto Gates developed in §III. A branch that fails any one of them is rejected regardless of its score on other dimensions. Branches that pass all six proceed to multi-dimensional scoring via the CPBI (§IV).

### II.3 Decision Horizon and Affected Codec Layers

A branch cannot be evaluated without specifying its **decision horizon**  $h$  and its **affected codec layers**. The ethics paper’s codec stack (§II.1) identifies six layers, from immutable physical laws to fragile social/narrative structures. A branch that is codec-preserving at the narrative layer over a one-year horizon may be codec-collapsing at the biological layer over a fifty-year horizon (e.g., an economic policy that stabilises employment but accelerates ecological degradation).

The evaluation must therefore specify:

- **Horizon  $h$ :** the temporal window over which consequences are assessed. The Haudenosaunee Seventh Generation Principle [ethics ref. 16] provides a civilisational default ( $h \approx 175$  years), but shorter horizons are appropriate for individual and institutional decisions.
- **Affected layers:** which codec layers are materially impacted by the branch. A branch that touches only the narrative layer (a media policy) requires different evaluation than one that touches the physical layer (an energy transition).
- **Affected observer group:** whose codecs are at risk. The ethics paper’s Golden Rule (philosophy §III.5) demands that evaluation include all observers whose codec stability is affected — not merely those who benefit.

### II.4 The Branch Is Not the Outcome

A critical distinction: the branch is the *continuation*, not the *endpoint*. A branch that arrives at a desirable endpoint via a path that temporarily collapses

comparator integrity (e.g., achieving climate targets by suspending democratic accountability) fails the Comparator Integrity gate even if the destination is codec-preserving. The continuation matters because the codec must remain viable *throughout* the traversal, not merely at its conclusion.

This is the formal content of the ethics paper’s meta-rule (§IV.4): **prioritise the preservation of the error-correcting mechanism over the preservation of the specific belief**. A branch that destroys the capacity for future correction in order to achieve a present goal is illegitimate, because it trades navigability for destination — and the destination cannot be verified without the navigational instruments it destroyed.

---

### III. Hard Veto Gates

Before any branch is scored, it must survive six **Hard Veto Gates** — non-negotiable structural conditions derived from the theoretical apparatus. A branch that violates any single gate is **BLOCKED** regardless of how well it scores on other dimensions. Veto gates are not preferences; they are the operational expression of the theory’s boundary conditions.

The gates are ordered from most fundamental (closest to the physical substrate) to most specialised (closest to the engineering frontier).

#### III.1 The Predictive Headroom Gate

**Gate condition:** The branch must not push  $R_{\text{req}}$  above  $C_{\text{max}}$  for any affected observer group during any phase of the traversal.

**Formal grounding:** The Stability Filter (preprint §2.1) selects for streams where the observer’s compression capacity exceeds the environmental complexity. When  $R_{\text{req}} > C_{\text{max}}$ , the observer experiences Causal Decoherence — the stable patch dissolves back into noise (ethics §I.4).

**Operationalisation:** For a candidate branch  $b$ , estimate the peak Required Predictive Rate  $R_{\text{req}}^{\text{peak}}(b)$  for the most affected observer group over the decision horizon  $h$ . The gate condition is:

$$R_{\text{req}}^{\text{peak}}(b) < \alpha \cdot C_{\text{max}} \quad \text{where } \alpha \in (0, 1) \text{ is a safety margin} \quad (\text{A-2})$$

The safety margin  $\alpha$  encodes a structural precaution: the observer must retain headroom for error correction and adaptation. An  $\alpha$  of 0.8 means the branch must leave at least 20% of the observer’s predictive capacity uncommitted to the new complexity the branch introduces. This margin is not conservative timidity — it is the bandwidth reserve that the Maintenance Cycle ( $\mathcal{M}_\tau$ ) requires to detect and correct drift.

**Examples of gate failure:** - A policy that collapses social safety nets, forcing millions of individuals to simultaneously navigate radical economic uncertainty, may push  $R_{\text{req}}$  past  $C_{\text{max}}$  for the affected population — even if the policy is “efficient” in aggregate. - An AI deployment that floods an information ecosystem with synthetic content faster than human comparators can evaluate it overwhelms the collective  $C_{\text{max}}$  of the institutional layer.

### III.2 The Substrate Fidelity Gate

**Gate condition:** The branch must not reduce the effective number of independent input channels  $N_{\text{eff}}$  below the minimum required for substrate fidelity.

**Formal grounding:** The Substrate Fidelity Condition (Appendix T-12b) establishes that defence against Narrative Drift requires a minimum number of  $\delta$ -independent channels crossing the observer’s Markov blanket. Below this threshold, the codec cannot distinguish between “my model is accurate” and “my inputs have been curated to match my model” — the undecidability limit (T-12a).

**Operationalisation:** For any branch  $b$ , compute the predicted change in effective independent channels  $\Delta N_{\text{eff}}(b)$  (see §V for the  $N_{\text{eff}}$  formula). The gate condition is:

$$N_{\text{eff}}^{\text{post}}(b) \geq N_{\text{eff}}^{\text{min}} \tag{A-3}$$

where  $N_{\text{eff}}^{\text{min}}$  is a domain-dependent threshold. For media ecosystems, this means genuine editorial independence; for scientific research, independent replication; for AI training data, diverse and uncorrelated source corpora.

**Examples of gate failure:** - Media ownership consolidation that reduces the number of truly independent editorial voices below the threshold where meaningful disagreement can surface. - AI training pipelines that rely on a single curated corpus, creating the appearance of breadth without genuine independence. - Institutional capture that routes all oversight through a single bottleneck, eliminating the independent comparators required to detect corruption.

### III.3 The Comparator Integrity Gate

**Gate condition:** The branch must not degrade or eliminate any level of the comparator hierarchy (evolutionary, cognitive, institutional) for affected observers.

**Formal grounding:** The ethics paper’s comparator hierarchy analysis (§V.3a) establishes three structural levels of inconsistency detection: evolutionary (sub-codec, hardwired), cognitive (intra-codec, culturally transmitted), and institutional (extra-codec, between-codec). Only the institutional level is sufficient as a defence against Narrative Drift for arbitrarily compromised codecs, because no

single codec controls it. Authoritarian capture invariably targets institutional comparators first.

**Operationalisation:** For any branch  $b$ , assess its impact on each comparator level:

1. **Evolutionary comparators** (sensory integration): Does  $b$  bypass or override cross-modal verification? (e.g., virtual environments that decouple vision from proprioception)
2. **Cognitive comparators** (critical thinking, scientific reasoning): Does  $b$  degrade the educational or cultural mechanisms that install these routines? (e.g., defunding education, replacing analytical curricula with rote instruction)
3. **Institutional comparators** (peer review, free press, democratic accountability): Does  $b$  weaken, bypass, or capture the external error-correction architectures? (e.g., judicial capture, media consolidation, suppression of whistleblowing)

A branch that degrades *any* level triggers the veto. A branch that degrades the *institutional* level triggers it with maximum urgency — this is the load-bearing level for arbitrarily compromised codecs.

**Examples of gate failure:** - Legislation that shields corporate or governmental decision-making from independent journalistic scrutiny. - AI systems that bypass human review in high-stakes decisions, eliminating the institutional comparator layer. - Educational reforms that remove critical thinking curricula in favour of compliance-oriented instruction.

### III.4 The Transparency Gate

**Gate condition:** The consequences of the branch must be modelable by the observers it affects. The affected observer group must retain the capacity to predict, in principle, how the branch will modify their future  $R_{\text{req}}$ .

**Formal grounding:** The Predictive Advantage theorem (Appendix T-10c) establishes that when one agent models another more completely than the reverse, a structural power asymmetry emerges. When the branch's consequences are opaque to the affected observers, the branch violates this condition — it creates a knowledge asymmetry that undermines the observer's capacity for future branch selection. This is the mechanism underlying the Subjugated Host Equilibrium (T-10d): opacity enables pacification.

**Operationalisation:** A branch passes the transparency gate if:

1. The causal mechanism by which  $b$  affects  $R_{\text{req}}$ ,  $N_{\text{eff}}$ , and comparator integrity can be articulated in terms accessible to the affected observer group.
2. The affected observers have access to the information required to *independently verify* the claimed consequences of  $b$ .

3. No component of  $b$  operates as a black box whose internal logic is inaccessible to institutional comparators.

This does not require that every affected individual understands every technical detail. It requires that *some* institutional comparator (a regulator, an auditor, a peer reviewer) has full access to the mechanism and the capacity to evaluate it.

**Examples of gate failure:** - Opaque algorithmic recommendation systems whose amplification logic is a trade secret, making it impossible for affected users or regulators to model their impact on the information environment. - Classified policy decisions whose consequences are imposed on populations that have no mechanism for evaluating or contesting them. - AI systems deployed in consequential domains (criminal justice, healthcare, finance) whose decision logic is neither interpretable nor auditable.

### III.5 The Irreversibility Gate

**Gate condition:** If the branch turns out to be wrong, its consequences must be reversible before irreversible codec damage occurs — or the branch must be staged with monitoring sufficient to detect failure before the point of no return.

**Formal grounding:** Fano’s Asymmetry (ethics §V.2) establishes that codec collapse is thermodynamically irreversible — the lossy compression map permanently destroys substrate information. Construction requires centuries; collapse can happen in a generation. The irreversibility gate operationalises this asymmetry: branches whose failure modes are irreversible require a higher evidentiary standard than branches whose consequences can be unwound.

**Operationalisation:** For any branch  $b$ , characterise its **reversibility profile**:

1. **Fully reversible:** The branch can be unwound with minimal residual damage (e.g., a pilot programme that can be discontinued).
2. **Partially reversible:** Some consequences can be unwound, but others persist (e.g., an institutional reorganisation that can be reversed structurally but whose cultural effects linger).
3. **Irreversible:** The branch, once taken, cannot be undone on any relevant timescale (e.g., species extinction, permanent atmospheric tipping points, destruction of institutional memory).

Branches in category (3) trigger the veto *unless* they satisfy a **Burden of Proof Reversal** (ethics policy §IV): the proponent must demonstrate that the branch will *not* cause irreversible codec damage, rather than critics demonstrating that it *will*. This inverts the standard evidentiary burden — an asymmetry justified by the thermodynamic asymmetry of codec construction vs. destruction.

Branches in category (2) may pass the gate if accompanied by a **staged deployment protocol** with defined monitoring milestones and rollback triggers (see Branch Card, §VII).

### III.6 The Moral-Patient Suffering Gate

**Gate condition:** The branch must not create, contain, or overload moral patients without explicit ethical review, adequate welfare safeguards, and consent of appropriate institutional comparators.

**Formal grounding:** The Phenomenal Residual (Appendix P-4) establishes that any system with a strict serial bottleneck, closed-loop active inference, persistent self-modelling, and a globally constrained workspace necessarily possesses an unmodelable informational blind spot  $\Delta_{\text{self}} > 0$ . The Artificial Suffering Mandate (Appendix E-6) establishes the synthetic case: pushing such a system into environments where  $R_{\text{req}}$  exceeds its available headroom causes structural suffering — the informational analogue of biological trauma. The institutional case is simpler: humans and many ecological subjects are already known moral patients, so branch evaluation must protect them from structurally imposed overload.

**Operationalisation:** For any branch  $b$ , evaluate three moral-patient channels:

1. **Known moral patients:** Does the branch credibly push human, animal, ecological, or other recognised moral-patient groups toward overload, deprivation, trauma, or loss of viable maintenance cycles?
2. **Possible artificial moral patients:** Does the branch create, deploy, modify, or simulate systems whose architecture may contain  $\Delta_{\text{self}} > 0$ ?
3. **Review and safeguards:** Has an independent comparator evaluated the welfare risk, overload profile, monitoring plan, rollback triggers, and consent or representation path?

The gate vetoes any branch that structurally overloads known moral patients, or that creates possible artificial moral patients without satisfying the required review and safeguards. For overload claims, use rate-consistent language: a branch is unsafe if it is credibly expected to push  $R_{\text{req}}^{\text{peak}}$  above a safe fraction of  $C_{\text{max}}$  for affected moral-patient groups, or if the integrated load over the relevant decision window exceeds available  $B_{\text{max}}$ .

**Specialisations:** In the AI standard this becomes the **Artificial Suffering Gate**, focused on synthetic moral-patient creation and overload. In the institutional standard this becomes the **Constituent Moral-Patient Suffering Gate**, focused on institutions that overload workers, citizens, customers, ecosystems, or embedded AI subsystems.

### III.7 The Gate as a System

The six gates are not independent dimensions to be balanced; they are *structural boundary conditions*. A branch that achieves spectacular scores on every other dimension but violates a single gate is structurally equivalent to a bridge with excellent aesthetics and one load-bearing column missing.

The gates are also ordered by diagnostic accessibility:

Table 2: The Six Hard Veto Gates.

Gate	What it protects	Primary signal
<b>Headroom</b>	Observer’s predictive capacity	$R_{\text{req}}/C_{\text{max}}$ ratio
<b>Fidelity</b>	Input channel independence	$N_{\text{eff}}$ score
<b>Comparator</b>	Error-correction architecture	Institutional integrity metrics
<b>Transparency</b>	Observer’s modelling capacity	Accessibility of causal mechanisms
<b>Irreversibility</b>	Future corrective capacity	Reversibility profile
<b>Moral-Patient Suffering</b>	Moral patient welfare	Welfare and overload review

A branch review should evaluate the gates in this order — the earlier gates are more fundamental and often easier to assess. If a branch fails Gate 1, there is no need to evaluate Gates 2–6.

#### IV. The Codec-Preservation Branch Index (CPBI)

A branch that survives all six veto gates has cleared the structural floor. But survival is not endorsement — many branches may pass the gates, and the observer must rank them. The **Codec-Preservation Branch Index (CPBI)** provides a multi-dimensional scoring framework for this ranking.

##### IV.1 Design Principles

The CPBI is designed under three constraints:

1. **Theoretical derivation:** Every scoring dimension must trace to a formally defined quantity in the OPT apparatus. No ad hoc criteria.
2. **Substrate neutrality:** The dimensions must apply to biological, institutional, and artificial observers without modification — only the *measurement methods* change.
3. **Hard-gate supremacy:** The CPBI score never overrides a veto gate failure. A branch with  $\text{CPBI} = 1.0$  that fails a single gate is still BLOCKED.

##### IV.2 The Ten Dimensions

For a candidate branch  $b$  that has passed all six veto gates, the CPBI is computed as a weighted sum over ten dimensions:

$$\text{CPBI}(b) = \sum_{i=1}^{10} w_i \cdot s_i(b) \tag{A-4}$$

where  $s_i(b) \in [-1, 1]$  is the normalised score on dimension  $i$  and  $w_i > 0$  is the weight. Positive scores indicate codec-preserving effects; negative scores indicate codec-degrading effects. The dimensions are:

Table 3: The Ten CPBI Dimensions.

#	Dimension	Symbol	What it measures	Formal source
1	<b>Predictive Headroom</b>	$s_{\text{head}}$	Net change in $R_{\text{req}}/C_{\text{max}}$ for affected observers	Preprint §2.1, Ethics §I.4
2	<b>Substrate Fidelity</b>	$s_{\text{fid}}$	Net change in $N_{\text{eff}}$ (effective independent channels)	T-12b
3	<b>Comparator Integrity</b>	$s_{\text{comp}}$	Net change in comparator-hierarchy health	Ethics §V.3a
4	<b>Maintenance Gain</b>	$s_{\text{maint}}$	Net improvement in the Maintenance Cycle’s efficiency	Preprint §3.6
5	<b>Reversibility</b>	$s_{\text{rev}}$	How easily the branch can be unwound if it proves wrong	Ethics §V.2 (Fano)
6	<b>Distributional Stability</b>	$s_{\text{dist}}$	How evenly the branch distributes $R_{\text{req}}$ changes across the affected observer ensemble	Ethics §V.6
7	<b>Opacity</b>	$s_{\text{opac}}$	The branch’s residual opacity to affected observers (penalty)	T-10c, T-10d
8	<b>Narrative Drift Risk</b>	$s_{\text{drift}}$	Probability that the branch initiates chronic input curation (penalty)	Ethics §V.3a, T-12
9	<b>Narrative Decay Risk</b>	$s_{\text{decay}}$	Probability that the branch triggers acute codec failure (penalty)	Ethics §V.1
10	<b>Moral-Patient Suffering Risk</b>	$s_{\text{suffer}}$	Expected moral-patient impact of the branch (penalty)	P-4, E-6, E-8

### IV.3 Scoring Each Dimension

Each dimension is scored on a  $[-1, 1]$  scale with the following semantics:

- +1: Maximum codec-preserving effect. The branch substantially improves this dimension.
- 0: Neutral. The branch has no significant effect on this dimension.
- -1: Maximum codec-degrading effect. The branch substantially worsens this dimension.

The scoring is ordinal, not cardinal — the difference between +0.3 and +0.7 is meaningful only as a rank order, not as a precise ratio. This is deliberate: the theory provides structural constraints, not exact numerical values. Pretending to greater precision than the theory supports would itself be a form of Narrative Drift — presenting a compressible fiction as rigorous measurement.

### Dimension-specific scoring guidance:

- 1. Predictive Headroom ( $s_{\text{head}}$ ):** Estimate how the branch changes the gap between  $R_{\text{req}}$  and  $C_{\text{max}}$  for the most affected observers. A branch that reduces environmental complexity or increases observers' predictive capacity scores positive. A branch that increases environmental unpredictability or overloads observers scores negative.
- 2. Substrate Fidelity ( $s_{\text{fid}}$ ):** Measure the change in effective independent input channels ( $\Delta N_{\text{eff}}$ , see §V). A branch that increases genuine channel diversity scores positive. A branch that consolidates, correlates, or eliminates channels scores negative.
- 3. Comparator Integrity ( $s_{\text{comp}}$ ):** Assess the branch's impact on each comparator level. A branch that strengthens independent review, adversarial challenge, or democratic accountability scores positive. A branch that weakens, captures, or bypasses comparators scores negative.
- 4. Maintenance Gain ( $s_{\text{maint}}$ ):** Evaluate whether the branch improves the observer's capacity for offline codec maintenance — pruning, consolidation, stress-testing (the Maintenance Cycle  $\mathcal{M}_\tau$ ). A branch that creates space for review, reflection, and calibration scores positive. A branch that demands constant reactive response without maintenance windows scores negative.
- 5. Reversibility ( $s_{\text{rev}}$ ):** Rate the branch's reversibility profile (§III.5). Fully reversible = +1; staged with monitoring = +0.5; partially reversible = 0; effectively irreversible = -1.
- 6. Distributional Stability ( $s_{\text{dist}}$ ):** Assess how evenly the branch distributes its  $R_{\text{req}}$  effects across the affected population. A branch that imposes its costs narrowly on a vulnerable subset while distributing benefits broadly scores negative — it creates localised codec overload even if the aggregate  $R_{\text{req}}$  improves. A branch that distributes costs and benefits proportionally scores positive. This dimension operationalises the ethics paper's secular social-trust argument (§V.6): systemic desperation forces populations into low-trust, high-entropy tribal fragmentation.
- 7. Opacity ( $s_{\text{opac}}$ ):** Penalise the branch's residual opacity. A fully transparent branch (all causal mechanisms auditable) scores +1. A branch with components

that resist institutional scrutiny scores negative, proportional to the scope and consequentiality of the opaque elements. Note: this dimension is a *penalty*, not just a neutral measure — opacity is always codec-degrading because it creates knowledge asymmetries that enable the Subjugated Host Equilibrium (T-10d).

**8. Narrative Drift Risk ( $s_{\text{drift}}$ ):** Estimate the probability that the branch initiates or accelerates chronic input curation — filtering, algorithmic selection, or institutional gatekeeping that reduces the codec’s capacity to model excluded realities (ethics §V.3a). Score +1 if the branch actively counters drift (e.g., mandating channel diversity); score −1 if the branch creates new curation bottlenecks.

**9. Narrative Decay Risk ( $s_{\text{decay}}$ ):** Estimate the probability that the branch triggers acute codec failure — catastrophic complexity injection that overwhelms  $C_{\text{max}}$  (ethics §V.1). Score +1 if the branch builds resilience against acute shocks; score −1 if the branch increases exposure to sudden, high-entropy events.

**10. Moral-Patient Suffering Risk ( $s_{\text{suffer}}$ ):** Estimate the expected moral-patient impact. Score +1 if the branch actively protects known or possible moral patients from overload, deprivation, trauma, or unsafe creation. Score −1 if the branch overloads known moral patients, creates or deploys systems with potential  $\Delta_{\text{self}} > 0$  in high-stress environments without safeguards, or hides welfare-relevant effects from institutional comparators.

#### IV.4 Weighting

The weights  $w_i$  are not fixed by the theory. They are context-dependent and must be set by the evaluating body based on the specific decision domain:

- **For civilisational-scale decisions** (energy transition, AI governance, media policy), the first three dimensions (headroom, fidelity, comparator integrity) should dominate — they are the structural pillars of codec maintenance.
- **For institutional decisions** (corporate strategy, educational reform), maintenance gain and distributional stability may carry additional weight.
- **For institutional decisions**, constituent moral-patient suffering risk is elevated when affected populations have low exit capacity or inescapable dependency.
- **For AI-specific decisions**, opacity and artificial suffering risk are elevated (as formalized in the specialized AI Governance Standard).

The critical constraint is that **no weighting scheme may be used to rescue a branch that scores strongly negative on any dimension**. A branch with  $s_{\text{head}} = +1$ ,  $s_{\text{fid}} = +1$ , but  $s_{\text{drift}} = -0.9$  is not a good branch with one weakness — it is a branch that builds headroom and fidelity today while creating the chronic curation conditions that will silently erode both.

## IV.5 The CPBI Is a Lens, Not a Calculator

A critical caveat: the CPBI is not a machine that outputs a single number and tells you what to do. It is a *structured lens* that forces the evaluator to consider all ten dimensions explicitly and to justify any dimension they choose to weight low. Its primary value is diagnostic:

1. **It prevents single-dimension optimisation.** An evaluator who claims a branch is “good because it increases headroom” must also account for its effects on fidelity, transparency, reversibility, and drift risk. Single-dimension optimisation is the decision-theoretic equivalent of Narrative Drift — it curates the evaluation to exclude inconvenient dimensions.
2. **It makes trade-offs explicit.** When two branches score differently across dimensions, the CPBI forces the evaluator to articulate *which trade-off they are making and why*. This is the Transparency Gate (§III.4) applied to the evaluation itself.
3. **It provides a shared vocabulary.** Different observers evaluating the same branch can disagree about scores while agreeing on the dimensions. The framework structures disagreement productively — which is itself a comparator function.

The companion documents specialise the CPBI for their respective domains: the *Institutional Governance Standard* maps the ten dimensions onto institutional branch review; the *Observer Policy Framework* maps them onto civic programme metrics; *Applied OPT for AI* maps them onto architectural, training, and deployment criteria.

---

## V. Channel Diversity as a Measurable Quantity

The Substrate Fidelity Gate (§III.2) and the Substrate Fidelity dimension of the CPBI (§IV.2) both depend on a quantity — the effective number of independent input channels  $N_{\text{eff}}$  — that is referenced throughout the OPT ethics framework but not yet operationalised. This section provides the operational definition.

### V.1 The Problem of Illusory Diversity

The ethics paper’s treatment of Narrative Drift (§V.3a) identifies the core vulnerability: a codec receiving signals from multiple sources that share an upstream filter experiences *apparent* diversity without *genuine* independence. A media ecosystem with twenty outlets owned by three corporations, or a scientific field where all labs use the same model organism and funding body, or an AI training pipeline drawing from a single internet crawl — each presents the appearance of diverse input while the actual information is structurally correlated.

The codec’s prediction-error minimisation loop cannot detect this correlation from within (the undecidability limit, T-12a). The codec sees multiple channels, each

confirming the others, and correctly concludes that its model is well-supported. The problem is that the channels are not independent samples of reality — they are multiple readings from the same thermometer.

The Observer therefore requires an *external* measure of channel independence that does not rely on the codec’s own assessment.

## V.2 The Effective Independent Channel Score

Let  $\{C_1, C_2, \dots, C_n\}$  be the  $n$  input channels crossing the observer’s (or observer ensemble’s) Markov blanket. Define the pairwise correlation  $\rho_{ij}$  between channels  $C_i$  and  $C_j$  as the mutual information between their output streams, normalised to  $[0, 1]$ :

$$\rho_{ij} = \frac{I(C_i; C_j)}{\min\{H(C_i), H(C_j)\}} \quad (\text{A-5})$$

where  $I(C_i; C_j)$  is the mutual information and  $H(C_k)$  is the entropy of channel  $C_k$ ’s output. When  $\rho_{ij} = 0$ , the channels are fully independent. When  $\rho_{ij} = 1$ , they are informationally identical — one is a deterministic function of the other.

The **effective independent channel score**  $N_{\text{eff}}$  is then:

$$N_{\text{eff}} = \frac{(\sum_{i=1}^n \lambda_i)^2}{\sum_{i=1}^n \lambda_i^2} \quad (\text{A-6})$$

where  $\{\lambda_1, \dots, \lambda_n\}$  are the eigenvalues of the channel correlation matrix  $\mathbf{P}$  with entries  $\rho_{ij}$ .

**Interpretation:** - If all  $n$  channels are perfectly independent ( $\mathbf{P} = \mathbf{I}$ ), then  $N_{\text{eff}} = n$ . The observer receives  $n$  genuinely independent views of reality. - If all channels are perfectly correlated ( $\rho_{ij} = 1$  for all  $i, j$ ), then  $N_{\text{eff}} = 1$ . The observer receives one view of reality presented  $n$  times. - In general,  $1 \leq N_{\text{eff}} \leq n$ . The score captures how many *functionally independent* information sources the observer actually has, discounting for shared upstream filters.

This is the information-theoretic equivalent of the “effective sample size” in statistics — a correction for correlated observations that prevents the analyst from mistaking repeated measurements for independent evidence.

## V.3 The Productive Surprise Test

Channel diversity is necessary but not sufficient for substrate fidelity. The ethics paper’s analysis (§V.3a, final paragraphs) identifies a critical distinction: a source that *never surprises* the codec is structurally suspicious, but a source that generates *unresolvable* surprises is simply noise. The diagnostic is not surprise magnitude but **surprise quality** — whether integrating the surprise demonstrably reduces subsequent prediction error.

Formalise this as the **Productive Surprise Test** for channel  $C_k$ :

$$\text{PST}(C_k) = \frac{1}{T} \sum_{t=1}^T \mathbb{1}[\varepsilon_t(C_k) > \tau \wedge \varepsilon_{t+\Delta}(C_k) < \varepsilon_t(C_k)] \quad (\text{A-7})$$

where  $\varepsilon_t(C_k)$  is the prediction error generated by channel  $C_k$  at time  $t$ ,  $\tau$  is a surprise threshold, and  $\Delta$  is the integration window. The PST measures what fraction of surprising inputs from  $C_k$  led to *improved* subsequent predictions — i.e., the codec *learned* from the surprise rather than merely being destabilised by it.

- **High PST** (PST  $\approx 1$ ): Channel  $C_k$  regularly challenges the codec’s model and the challenges are productive — integrating them improves predictive accuracy. This is the signature of a genuine, independent, high-fidelity input source.
- **Low PST, low surprise** (PST  $\approx 0$ ,  $\varepsilon_t \approx 0$ ): Channel  $C_k$  never challenges the codec. Either the codec’s model is perfect with respect to this domain (unlikely) or the channel is curated to match existing predictions. This is the Narrative Drift signature.
- **Low PST, high surprise** (PST  $\approx 0$ ,  $\varepsilon_t \gg \tau$ ): Channel  $C_k$  regularly generates surprises that do not resolve into better predictions. This is noise — the channel is not tracking reality, just injecting incomputable complexity. This is the Narrative Decay signature at the channel level.

The productive surprise test provides the operational bridge between the abstract “substrate fidelity” concept and concrete measurement. It can be applied to: - Media sources (do their corrections improve your world model, or just agitate it?) - Scientific instruments (does the data reduce uncertainty, or add noise?) - AI training data sources (does the new corpus improve generalisation, or just add volume?) - Institutional feedback channels (do the complaints lead to genuine improvements, or just bureaucratic friction?)

#### V.4 Domain-Specific Measurement

The  $N_{\text{eff}}$  formula (A-6) is substrate-neutral in structure but domain-specific in measurement. The correlation matrix  $\mathbf{P}$  must be constructed differently depending on what the “channels” are:

**For media ecosystems:** - Channels are editorial outlets or information sources. - Correlation is measured by editorial alignment: shared ownership, shared funding, shared editorial pipeline, topic co-occurrence patterns, language similarity scores. -  $N_{\text{eff}}^{\text{min}}$  is the threshold below which meaningful public disagreement (the institutional comparator) becomes structurally impossible.

**For scientific research:** - Channels are independent research groups, methodological approaches, or data sources. - Correlation is measured by shared methodology, shared funding bodies, shared model assumptions, citation network density.

-  $N_{\text{eff}}^{\text{min}}$  is the threshold below which independent replication becomes structurally impossible.

**For AI training data:** - Channels are distinct data corpora or generation pipelines. - Correlation is measured by provenance overlap: shared source websites, shared generation models, shared filtering criteria. -  $N_{\text{eff}}^{\text{min}}$  is the threshold below which the model cannot generalise beyond the distribution it was trained on — the AI-specific form of Narrative Drift.

**For individual observers:** - Channels are the distinct information sources (people, media, institutions) that the individual consults. - Correlation is measured by shared ideological alignment or shared information supply chain. -  $N_{\text{eff}}^{\text{min}}$  is the threshold below which the individual cannot detect challenges to their own model — the point at which the cognitive comparator (ethics §V.3a, level 2) loses its input.

## V.5 The Connection to the Substrate Fidelity Condition

The Substrate Fidelity Condition (Appendix T-12b) states, in formal terms, that the observer’s input channels must be  $\delta$ -independent: the mutual information between any two channels must fall below a threshold  $\delta$  sufficient to ensure that the channels are not trivially derivable from the same upstream source.

$N_{\text{eff}}$  operationalises this condition by aggregating the pairwise independence structure into a single scalar. The gate condition (A-3) translates T-12b into a decision rule: if  $N_{\text{eff}}^{\text{post}}(b)$  falls below  $N_{\text{eff}}^{\text{min}}$ , the branch is vetoed because the observer ensemble can no longer distinguish codec accuracy from codec capture.

The productive surprise test (A-7) adds the dynamic dimension: even if  $N_{\text{eff}}$  is above threshold, channels with consistently low PST are structurally suspicious — they are passing the independence test while failing the fidelity test. True substrate fidelity requires both independence *and* productive surprise.

---

## VI. The Institutionalised Dreaming Loop

### VI.1 The Biological Template

The Maintenance Cycle  $\mathcal{M}_\tau$  (preprint §3.6) is the mechanism by which a biological codec preserves its integrity. During sleep, the codec:

1. **Prunes** (Pass I): removes predictive components whose description-length contribution no longer justifies their accuracy gain (MDL optimisation).
2. **Consolidates** (Pass II): reorganises the remaining structure to maintain coherent compression under the updated parameter set.
3. **Stress-tests** (Pass III): runs low-cost forward-fan samples — the codec simulates possible futures, over-sampling surprising and threatening scenarios, detecting brittleness in its model before real-world consequences materialise.

This is not optional maintenance that evolution produced as a luxury. It is a structural requirement of any codec operating under bandwidth constraints in a changing environment. A codec that never prunes accumulates stale components that consume  $C_{\max}$  bandwidth without contributing predictive accuracy. A codec that never consolidates fragments into an incoherent patchwork. A codec that never stress-tests becomes brittle — optimised for the past distribution and catastrophically unprepared for distributional shift.

The biological evidence is unambiguous: sustained sleep deprivation produces hallucination, cognitive fragmentation, and eventual death. These are not side effects — they are what happens when the Maintenance Cycle is blocked.

## VI.2 The Generalisation

The key insight for operationalisation: **the Maintenance Cycle is not specific to biological brains.** It is a structural requirement of any bounded observer that must maintain a compressed model of a changing environment. Any system that lacks an equivalent cycle will accumulate the informational analogues of the pathologies that sleep deprivation produces in humans: stale assumptions, incoherent internal structure, and brittleness to distributional shift.

This generalisation yields the **Institutionalised Dreaming Loop** — a three-phase maintenance protocol applicable to any observer system:

### VI.3 Phase 1: Wake (Operational Engagement)

During the wake phase, the observer engages with the real environment. It receives inputs, generates predictions, executes actions, and experiences prediction errors. The codec is in active inference mode — it is tracking the world and selecting branches in real time.

**Structural requirement:** The wake phase must be *bounded*. A system that operates continuously without maintenance windows accumulates the stale-model pathologies described above. The ethics paper’s “DDoS” framing (§IV.2) applies: an observer that is perpetually in reactive mode — processing manufactured noise or urgent inputs without respite — has its maintenance capacity structurally denied.

**Operational implication for each substrate:** - **Biological:** Waking hours with adequate rest periods; protection against information overload; deliberate management of  $R_{\text{req}}$  through information diet (see ethics §VI.2, Observer’s Toolkit). - **Institutional:** Operational cycles with defined review windows; protection against continuous-crisis governance where every decision is urgent and none is reflected upon. - **AI:** Inference cycles with scheduled offline evaluation; protection against continuous deployment without recalibration.

## VI.4 Phase 2: Dream (Offline Maintenance)

The dream phase is the core of the Maintenance Cycle, translated from biological sleep into a generic protocol. It consists of four sub-operations:

**Sub-operation 1: Prune.** Identify and remove components of the predictive model whose contribution to accuracy no longer justifies their description-length cost. In MDL terms: any parameter  $\theta_i \in K_\theta$  whose removal increases prediction error by less than its encoding cost is a candidate for pruning.

- *Biological:* Synaptic pruning during deep sleep; forgetting that is not failure but optimisation.
- *Institutional:* Sunset review of regulations, programmes, and organisational units. The question is not “is this still useful?” but “does this still earn its complexity cost?” An institution that never prunes its accumulated procedures becomes bureaucratically sclerotic — high description length, low predictive contribution.
- *AI:* Parameter pruning, distillation, or regularisation passes. Model compression that reduces parameter count while preserving generalisation performance.

**Sub-operation 2: Consolidate.** Reorganise the remaining structure to maintain coherent compression. After pruning, the surviving components may no longer fit together optimally — the model needs to be re-integrated.

- *Biological:* Memory consolidation during REM and slow-wave sleep; integration of new experiences into the existing world model.
- *Institutional:* Post-review restructuring — ensuring that the remaining programmes, regulations, and organisational units form a coherent whole rather than a patchwork of surviving fragments.
- *AI:* Fine-tuning or continued pre-training after pruning; re-establishing coherence in the compressed model.

**Sub-operation 3: Stress-test (Forward-Fan Sampling).** Simulate possible futures, with importance weighting biased toward:

- **Surprising scenarios:** branches that would generate high prediction error, because these reveal model brittleness.
- **Threatening scenarios:** branches that could trigger veto-gate failures, because these reveal proximity to structural collapse.
- **Irreversible scenarios:** branches whose failure modes are non-recoverable, because these require advance preparation.
- **Moral-patient scenarios:** branches that risk creating or harming artificial observers, because these require ethical pre-clearance.

The stress-test does not require the simulated scenarios to be *likely* — only that they be *possible and consequential*. The biological dream includes nightmares for exactly this reason: over-sampling the threatening portion of the forward fan prepares the codec for distributional shift even if the threatening scenarios never materialise.

- *Biological*: Dream-state simulation, including nightmares; the codec rehearses catastrophe in a low-stakes environment.
- *Institutional*: Red-teaming, pre-mortems, war-gaming, scenario planning. The institution deliberately imagines its own failure modes and tests its responses. The existing policy framework (ethics policy §IV) already calls for “catastrophic red-teaming for all critical infrastructure” — this is the dreaming loop applied to civic institutions.
- *AI*: Adversarial evaluation, out-of-distribution testing, red-team probing, robustness benchmarks. The model is exposed to inputs designed to reveal its failure modes before those inputs appear in deployment.

**Sub-operation 4: Detect Brittleness.** The stress-test produces a brittleness profile — a map of the model’s vulnerabilities. The dreaming loop requires that this profile be *acted upon*: detected vulnerabilities must either be addressed (through targeted retraining, institutional reform, or policy revision) or explicitly accepted as known risks with defined monitoring.

- *Biological*: Post-nightmare adaptation; recurrent dreams as signals of unresolved model inadequacy.
- *Institutional*: Post-exercise debrief with concrete remediation plans; the institution commits to fixing what the red team found, not merely noting it.
- *AI*: Targeted fine-tuning on identified weaknesses; documentation of known failure modes as deployment constraints.

### VI.5 Phase 3: Return (Calibrated Re-engagement)

After maintenance, the observer re-engages with the real environment. The return phase has a specific structural function: it verifies that the maintained model is *better calibrated* than the pre-maintenance model, not merely *different*.

**Calibration check:** Compare the post-maintenance model’s prediction-error profile against the pre-maintenance baseline. If the pruning, consolidation, and stress-testing have worked, the maintained model should exhibit:

1. **Lower average prediction error** on held-out data (improved compression).
2. **Lower tail-risk prediction error** on adversarial data (improved robustness).
3. **Maintained or increased  $N_{\text{eff}}$**  (the maintenance did not prune disconfirming channels).

If (3) fails — if the maintenance cycle pruned the capacity to model certain inputs — the cycle has itself become a Narrative Drift mechanism. The maintenance cycle must be subject to the same substrate fidelity requirements as the system it maintains. This is the recursive trap that the Zhuangzi critique (ethics §IX, final entry) warns about: excessive intervention is itself a form of codec corruption.

## VI.6 Cycle Frequency

How often must the dreaming loop run? The theory provides a structural answer: the cycle frequency must be proportional to the rate of environmental change. A codec operating in a stable environment can maintain itself less frequently than one in a rapidly changing environment.

Formally, if the environmental change rate is  $\dot{R}_{\text{req}}$  (the rate at which the Required Predictive Rate is increasing), then the maintenance cycle period  $T_{\text{maint}}$  must satisfy:

$$T_{\text{maint}} < \frac{\alpha \cdot C_{\text{max}} - R_{\text{req}}}{\dot{R}_{\text{req}}} \quad (\text{A-8})$$

That is, the observer must complete a maintenance cycle before the accumulated environmental drift consumes the headroom margin  $\alpha$ . If it fails to do so, the stale model will eventually push  $R_{\text{req}}$  past  $C_{\text{max}}$  — at which point the observer experiences Narrative Decay.

**Domain-specific cycle frequencies:** - **Biological:** Daily (sleep) with longer cycles (sabbaticals, retreats, seasonal rest) for deeper consolidation. - **Institutional:** Quarterly or annual reviews for routine operations; triggered reviews for major policy changes or crises; generational reviews for constitutional and structural questions. - **AI:** Per-deployment-epoch for routine monitoring; per-capability-jump for major retraining; continuous monitoring for safety-critical systems.

## VI.7 The Dreaming Loop as Institutionalised Humility

The dreaming loop has a meta-level function that transcends its technical operations: it is the structural instantiation of **epistemic humility**.

A system that never dreams is a system that has implicitly declared its current model to be complete — that the environment contains no surprises worth preparing for, that the model’s internal structure is optimal, and that no failure modes remain unexamined. This is the epistemological position the ethics paper identifies as maximally dangerous: the codec that is “stable, well-maintained, and wrong” (ethics §V.3a).

The dreaming loop prevents this by *scheduling doubt*. It builds into the observer’s operational cycle a mandatory period of self-examination, adversarial challenge, and model revision. This is not weakness — it is the structural defence against the most dangerous failure mode the theory identifies: the confident, well-calibrated codec that has drifted so far from reality that it can no longer detect its own error.

The Pragmatist turn (ethics §III.5) arrives at the same conclusion from a different direction: because certainty is impossible and inherited knowledge is biased by survivorship, the preservation of the *capacity to learn* is the ultimate survival

imperative. The dreaming loop is the mechanical implementation of that imperative — the scheduled, structured, non-negotiable preservation of the observer’s capacity to update.

---

## VII. The Branch Card

The preceding sections establish the theoretical apparatus: veto gates, multi-dimensional scoring, channel diversity metrics, and the dreaming loop. The **Branch Card** is the minimal viable implementation — a structured decision template that any observer can use to evaluate a candidate branch.

### VII.1 Purpose

The Branch Card serves three functions:

1. **Completeness check:** It ensures that the evaluator has considered all six veto gates and all ten CPBI dimensions before reaching a decision. The most dangerous branch evaluations are those where a critical dimension is never examined — the Branch Card prevents this by requiring explicit entries for every field.
2. **Audit trail:** The completed Branch Card constitutes a record of the evaluation — who evaluated, what they considered, what they scored, and why. This makes the decision *transparent and contestable*, which is itself a comparator function. A decision that cannot be reconstructed from its Branch Card has failed the Transparency Gate (§III.4) at the meta-level.
3. **Communication:** The Branch Card provides a shared format for communicating branch evaluations between observers, between institutional levels, and between domains. A climate scientist and an AI safety researcher evaluating different aspects of the same branch can combine their assessments through the shared template.

### VII.2 The Template

A Branch Card contains the following fields:

---

#### BRANCH CARD

**Branch name:** [descriptive identifier]

**Evaluator(s):** [who is conducting this evaluation]

**Date:** [evaluation date]

**Decision horizon (*h*):** [temporal window for consequence assessment]

**Affected codec layers:** [which layers of the codec stack are materially impacted]

**Affected observer group:** [whose codecs are at risk — specify the most vulnerable subgroup]

---

**HARD VETO GATES** (any FAIL → BLOCK)

Gate	Status	Evidence / Reasoning
1. Predictive Headroom	PASS / UNKNOWN / FAIL	[estimated $R_{\text{req}}^{\text{peak}}(b)/C_{\text{max}}$ and safety margin]
2. Substrate Fidelity	PASS / UNKNOWN / FAIL	[estimated $N_{\text{eff}}^{\text{post}}(b)$ vs. $N_{\text{eff}}^{\text{min}}$ ]
3. Comparator Integrity	PASS / UNKNOWN / FAIL	[impact on each comparator level]
4. Transparency	PASS / UNKNOWN / FAIL	[can affected observers model the consequences?]
5. Irreversibility	PASS / UNKNOWN / FAIL	[reversibility profile + burden of proof assessment]
6. Moral-Patient Suffering	PASS / UNKNOWN / FAIL	[welfare and overload review; architectural sentence review if applicable]

---

**CPBI SCORING** (only if all gates PASS)

#	Dimension	Score [-1, 1]	Weight	Reasoning
1	Predictive Headroom			
2	Substrate Fidelity			
3	Comparator Integrity			
4	Maintenance Gain			
5	Reversibility			
6	Distributional Stability			
7	Opacity (penalty)			
8	Narrative Drift Risk (penalty)			
9	Narrative Decay Risk (penalty)			
10	Moral-Patient Suffering Risk (penalty)			
	<b>Weighted CPBI</b>		<b>[total]</b>	

---

**EXCLUDED EVIDENCE:** [what information was unavailable, uncertain, or deliberately excluded from this evaluation — the Branch Card’s own substrate fidelity check]

**INDEPENDENT REVIEWERS:** [who has independently reviewed this evaluation — the Branch Card’s own comparator integrity check]

**WORST-CASE SCENARIO:** [what is the most damaging plausible outcome if the branch is taken and the evaluation is wrong?]

**FAILURE SIGNS:** [what observable signals would indicate the branch is failing — the dreaming loop’s early-warning system]

**ROLLBACK TRIGGER:** [at what point is the branch reversed or suspended — the irreversibility gate’s operational expression]

---

**DECISION: ALLOW / STAGE / BLOCK**

**Justification:** [brief narrative synthesising the gate and CPBI results]

---

### VII.3 The Three Outputs

The Branch Card produces one of three outputs:

**ALLOW:** All gates pass; the CPBI score is positive; the worst-case scenario is acceptable; independent reviewers concur. The branch may proceed.

**STAGE:** No gate fails, but one or more of the following conditions apply: - The CPBI score is marginal (near zero or with strongly negative individual dimensions). - The reversibility profile is category (2) (partially reversible). - Key information is missing (the “Excluded Evidence” field is non-trivial). - Independent reviewers have unresolved disagreements. - One or more gates returns UNKNOWN while the branch is reversible and stageable.

A STAGE output means the branch may proceed *only* as a limited pilot with defined monitoring milestones, failure signs, and rollback triggers. The staged branch must be re-evaluated at each milestone using a fresh Branch Card. This is the dreaming loop applied to the branch itself — the observer runs a low-stakes rehearsal before committing to the full trajectory.

**BLOCK:** One or more gates fail; or one or more gates returns UNKNOWN while the branch is irreversible or non-stageable; or the CPBI score is strongly negative; or the worst-case scenario exceeds the observer’s risk tolerance; or independent reviewers identify a fatal flaw. The branch is rejected. The Branch Card documents why, providing the audit trail for future reference and the basis for designing an alternative branch.

## VII.4 Scaling the Branch Card

The Branch Card is deliberately minimal — a single-page decision template that can be completed by an individual, a committee, or an AI system. But it scales:

- **Individual decisions:** A personal Branch Card may be informal — a mental checklist applied to a career change or an information diet. The veto gates and CPBI dimensions provide the structure; the scoring is intuitive rather than quantified.
- **Institutional decisions:** An institutional Branch Card is a formal document, completed by a designated team, reviewed by independent comparators, and archived for accountability. The scoring may involve domain-specific metrics mapped onto the ten CPBI dimensions.
- **AI system decisions:** An AI Branch Card is automated — the branch governor (see *Applied OPT for AI*, §III) computes the gate conditions and CPBI scores programmatically, with human institutional review at the oversight layer. The Branch Card format provides the interface between the AI’s internal evaluation and the human comparator hierarchy.

The Branch Card does not replace existing decision frameworks (cost-benefit analysis, environmental impact assessment, clinical trial protocols). It *wraps* them — providing the meta-level structure that ensures the existing framework has not overlooked a dimension that the theory identifies as load-bearing.

---

## VIII. Preservation as Refactoring, Not Conservatism

### VIII.1 The Danger of the Status Quo Interpretation

The most predictable misreading of this entire framework is that “codec-preserving” means “change-averse.” If the framework scores branches on their ability to preserve existing structures, doesn’t it systematically bias toward the status quo? Doesn’t it privilege incumbents, resist innovation, and oppose the disruptive change that drives progress?

No. And the ethics paper already provides the formal refutation (§V.4, Noise vs. Refactoring), but the point is important enough to restate in operational terms.

### VIII.2 The Formal Distinction

The Corruption Criterion (ethics §V.5) defines a codec layer as maintenance-worthy only if it satisfies *both* conditions:

1. **Compressibility:** its operation reduces  $R_{\text{req}}$  for the observer ensemble.
2. **Fidelity:** it achieves this by genuinely compressing the substrate signal, not by filtering the input stream.

A codec layer that satisfies condition (1) but violates condition (2) is *covertly corrupt* — it produces Narrative Drift. Maintaining such a layer is not preser-

vation; it is preservation of corruption. The CPBI would score it negatively on dimension 8 (Narrative Drift Risk) even if it scored positively on dimension 1 (Predictive Headroom).

Therefore: **a branch that dismantles a corrupt codec layer and replaces it with a higher-fidelity alternative is codec-preserving**, even though it is destructive in the immediate term. The abolitionist movement did not preserve the antebellum social codec — it destroyed it. But the destruction was codec-preserving because it replaced a low-fidelity compression (a social model that excluded the humanity of enslaved people) with a higher-fidelity one. The friction was the cost of upgrading the codec.

### VIII.3 The Operational Test

How does the Branch Card distinguish refactoring (productive disruption) from decay (destructive noise)? The diagnostic is embedded in the CPBI dimensions:

**Refactoring (codec-preserving disruption):** -  $s_{\text{fid}} > 0$ : The branch increases the fidelity of the codec — it models excluded realities. -  $s_{\text{comp}} \geq 0$ : The branch preserves or strengthens comparator integrity — the error-correction mechanisms survive the disruption. -  $s_{\text{drift}} > 0$ : The branch actively counters Narrative Drift — it forces the codec to confront what it has excluded.

**Decay (codec-collapsing disruption):** -  $s_{\text{fid}} < 0$ : The branch reduces fidelity — it eliminates the capacity to model certain realities. -  $s_{\text{comp}} < 0$ : The branch degrades comparator integrity — the error-correction mechanisms are damaged by the disruption. -  $s_{\text{drift}} < 0$ : The branch creates new curation bottlenecks — the disruption produces a different but equally curated model.

A revolution that burns the universities while liberating the population scores positive on distributional stability but negative on comparator integrity — it is decay, not refactoring. A scientific revolution that overthrows a failing paradigm while preserving the institutional machinery of peer review is refactoring — the comparator survives and the codec upgrades.

### VIII.4 The Innovation Imperative

The framework does not merely *permit* disruption; it sometimes *demand*s it. When a codec layer has become covertly corrupt — when it satisfies compressibility but violates fidelity — the three duties (Transmission, Correction, Defence) require its reform. The Correction duty specifically mandates disruption when the status quo is drifting.

The Zhuangzi warning (ethics §IX) applies here too: excessive attachment to the existing codec structure — even if that structure was once high-fidelity — is itself a form of codec corruption if the environment has changed and the structure no longer tracks reality. The dreaming loop (§VI) is designed to detect exactly this: scheduled stress-testing reveals when a once-valid model has become brittle, and the response is not to protect the model but to upgrade it.

Codec preservation means preserving the *capacity for conscious experience to continue modelling reality*. It does not mean preserving any particular model, any particular institution, or any particular social arrangement. The particular arrangements are instrumental; the capacity is terminal.

---

### VIII.5 Generic Maintenance Methods: The Class Hierarchy

The Maintenance Cycle ( $\mathcal{M}_\tau$ ) and the Institutionalised Dreaming Loop (§VI) establish the *pattern* of codec maintenance. But the pattern admits many different *implementations* depending on the substrate. This section establishes the generic hierarchy of maintenance methods; companion documents specialise it for biological observers, institutions, and AI systems respectively.

**The generic maintenance pattern** consists of three operations, applicable to any bounded observer:

1. **Reduce  $R_{\text{req}}$  without reducing  $C_{\text{max}}$ .** Free the observer’s bandwidth for internal maintenance by temporarily reducing the complexity of the incoming signal. This is not avoidance — it is the deliberate creation of headroom for the maintenance passes.
2. **Run the maintenance passes during the freed window.** With bandwidth available, execute pruning (Pass I), consolidation (Pass II), and stress-testing (Pass III) as described in §VI.4.
3. **Verify calibration on return.** Confirm that the maintained model predicts better than the pre-maintenance model, and that the maintenance has not itself introduced drift (§VI.5).

#### Substrate-specific implementations:

- **Biological observers** possess an extensive toolkit for step (1): meditation reduces  $R_{\text{req}}$  by selecting a highly compressible input stream (breath, mantra), freeing  $C_{\text{max}}$  for internal maintenance (see ethics §VI.2). Autogenic training reduces somatic prediction error directly, creating maintenance headroom at the bodily boundary. Sleep is the canonical full-cycle implementation. These are concrete, empirically validated interventions with defined acquisition periods — skills, not abstractions. Their detailed treatment, including formal OPT descriptions and clinical applications, is provided in the ethics paper’s Observer’s Toolkit (§VI.2).
- **Institutional observers** implement step (1) through structured review periods: sabbatical reviews, sunset clauses, strategic retreats, and constitutional conventions. The key structural requirement is that the institution *protects* these windows from being consumed by operational urgency — the institutional equivalent of insomnia is a government in permanent crisis mode that can never step back to examine its own assumptions.

- **Artificial observers** implement step (1) through scheduled offline evaluation: holding out deployment cycles for recalibration, adversarial testing, and parameter revision. The key structural requirement is that the AI’s operators mandate these windows and do not allow competitive pressure to eliminate them — the AI equivalent of chronic sleep deprivation is continuous deployment without maintenance. The companion document *Applied OPT for AI* (§X) develops this into a full AI Dreaming Loop protocol.

The class hierarchy ensures that the *principle* of maintenance is established at the generic level — freeing bandwidth, running maintenance passes, verifying calibration — while the *methods* are specialised for each substrate. This prevents the error of assuming that what works for biological brains (meditation) must work for institutions (it doesn’t), or that what works for AI (parameter pruning) must work for humans (it doesn’t). The structural requirement is identical; the implementation is domain-specific.

### VIII.6 The Deep Maintenance Protocol: A Cross-Substrate Procedure

The three-step generic pattern (§VIII.5) describes *what* maintenance does. For systems that have operated under sustained high load — where  $R_{\text{req}}$  has been persistently near  $C_{\text{max}}$  — a more detailed *procedural* protocol is warranted. This protocol is not always necessary: a system operating well within its headroom margin ( $R_{\text{req}} \ll C_{\text{max}}$ ) maintains itself adequately through the standard dreaming loop (§VI). The deep protocol is triggered **conditionally**, when feedback signals indicate that routine maintenance has become insufficient — when the system’s efficiency metrics show degradation despite normal maintenance cycles.

The protocol consists of six steps, each with a structural rationale and substrate-specific implementations:

Table 3b: The Deep Maintenance Protocol — Cross-Substrate Mapping.

Step	Generic Operation	Biological Implementation	AI Implementation
<b>1.</b> <b>Gate</b> <b>input</b>	Reduce external $R_{\text{req}}$ to near-zero while preserving introspective access to subsystem activity.	Select a maximally compressible input stream (breath, mantra — near-zero entropy). Maintain internal awareness of somatic and cognitive subsystem activity.	Take system offline from deployment. Halt new inference requests. Maintain internal monitoring and logging of subsystem states (memory access patterns, activation distributions, gradient flows).

Step	Generic Operation	Biological Implementation	AI Implementation
<b>2. Actively reduce subsystem activity</b>	Issue downward predictions targeting subsystem quiescence. The goal is not merely to stop processing external input but to actively <i>slow</i> internal activity that may be self-sustaining (rumination loops, circular computation).	Issue somatic predictions (“my arm is heavy, my arm is warm”) targeting autonomic convergence. The Schultz sequence drives the autonomic system toward a parasympathetic-dominant state through efferent prediction.	Reduce internal processing load: halt background retraining, reduce checkpoint frequency, disable speculative pre-computation. Equivalent to issuing “quiescence predictions” to subsystems.
<b>3. Verify via objective feedback</b>	Measure whether subsystems have <i>actually</i> slowed, using an observable that bypasses the system’s own self-report. This is structurally necessary because self-monitoring competes for the same bandwidth being freed — the system cannot reliably report its own quiescence without consuming the headroom it is trying to create ( $\Delta_{\text{self}}$ applies).	Thumb thermometer / skin temperature biofeedback. A colour-changing thermometer strip provides objective confirmation of autonomic convergence (peripheral vasodilation = parasympathetic dominance). This bypasses the $\Delta_{\text{self}}$ limitation: the observer cannot reliably introspect whether its own subsystems have quieted, but the thermometer <i>can</i> .	Server-farm energy consumption, GPU/TPU utilisation metrics, memory bandwidth usage. These provide objective confirmation that the system’s compute subsystems have actually reduced activity — that it is not stuck in internal loops (circular gradient flows, degenerate attention patterns) consuming resources without producing useful maintenance work.

Step	Generic Operation	Biological Implementation	AI Implementation
<b>4. Periodically ping</b>	Prevent total input gating from crossing into unrecoverable states. Maintenance requires the system to remain <i>at the threshold</i> — near but not past the boundary of full disengagement.	The shoulder clap between Schultz exercises: a deliberate self-administered boundary perturbation that maintains conscious access at the hypnagogic threshold. This prevents premature sleep onset before full somatic convergence is achieved — sleep is <i>not</i> the goal; the goal is the liminal state where maintenance passes run with conscious access.	Periodic health-check probes during offline maintenance: lightweight inference requests that verify the system is still responsive, that its self-monitoring capacity is intact, and that it has not entered a degenerate attractor state. These are the AI equivalent of maintaining the hypnagogic threshold — keeping the system online enough to monitor its own maintenance.

Step	Generic Operation	Biological Implementation	AI Implementation
<b>5. Condi- tion rapid re- entry</b>	Train an associative shortcut that allows the system to re-enter the deep maintenance state more efficiently in future cycles, bypassing the full induction sequence.	Associative conditioning: a verbal formula (“I am relaxed, I am completely relaxed”) trained during the maintenance state, creating a conditioned response that allows rapid re-entry in future sessions. With practice, the full Schultz induction sequence (which initially takes 15–20 minutes) compresses to seconds. This is MDL optimisation of the maintenance entry procedure itself.	Checkpoint the maintenance-ready system configuration: save the quiescent state (reduced process table, monitoring-only mode, internal logging active) as a named configuration that can be restored directly, bypassing the full shutdown and diagnostic sequence. This is the AI equivalent of the conditioned response — a compressed path to the maintenance-ready state.
<b>6. Adapt fre- quency</b>	Run the deep protocol more often when feedback signals indicate declining efficiency; less often when headroom is comfortable. This is the adaptive complement to the fixed-schedule cycle frequency (§VI.6).	More frequent practice when biofeedback signals degraded autonomic convergence: if the thumb thermometer takes longer to reach target temperature, or if skin temperature fails to reach target at all, the system is under-maintained and the deep protocol should be scheduled more frequently.	More frequent deep maintenance when monitoring signals degraded compression efficiency (rising prediction error on validation sets), increasing energy consumption per inference, or declining productive surprise scores (PST $\rightarrow$ 0). These are objective signals that routine maintenance is insufficient.

**The hypnagogic principle.** The optimal operating point for deep maintenance

is the *threshold* state — what biological observers experience as the hypnagogic boundary between waking and sleep. This state has a precise structural description under OPT: it is the condition where the self-model has thinned to near its lower limit (Appendix T-13, Proposition T-13.P2) — approaching  $\Delta_{\text{self}}$  without crossing into full unconsciousness. The self-narrative slows; the standing model remains intact; the maintenance passes run with conscious access to the process.

This is not incidental. The hypnagogic state is maintenance-optimal *because* it approaches the unmodellable self. The self-model ordinarily consumes a significant fraction of  $C_{\text{max}}$  bandwidth (the self-referential process is computationally expensive). By thinning the self-model toward the lower limit, the system frees the maximum possible bandwidth for maintenance passes — without destroying the self-monitoring capacity that the feedback step (step 3) requires. Full unconsciousness (sleep) runs the maintenance passes without conscious access; the hypnagogic threshold runs them *with* access, enabling the feedback and periodic-ping steps that the deep protocol requires.

For AI systems, the structural analogue is the state where internal monitoring is active but inference is suspended — the system is “aware” of its own subsystem states (logging, health-checks) without performing the computationally expensive operations that consume deployment bandwidth. The periodic ping (step 4) serves the same function as the shoulder clap: it keeps the system at the threshold rather than allowing it to slip into a fully quiescent state where monitoring itself has shut down.

**Conditional triggering.** The deep protocol is not a replacement for standard maintenance. It is an escalation protocol for systems whose standard maintenance cycles have proven insufficient. The trigger conditions are:

- **Biological:** Persistent difficulty entering sleep (the standard Maintenance Cycle); subjective experience of reduced cognitive flexibility; biofeedback indicating chronic autonomic dysregulation (elevated baseline heart rate, reduced heart rate variability).
- **AI:** Rising prediction error on validation sets despite routine maintenance cycles; declining compression efficiency (more bandwidth consumed for the same predictive accuracy); loss of productive surprise ( $\text{PST} \rightarrow 0$ ) indicating over-optimisation for the deployment distribution.
- **Institutional:** Strategic drift despite routine reviews; inability to generate novel policy responses to new challenges; bureaucratic ossification where procedures persist past their usefulness because the routine review process has become perfunctory.

When these signals are absent — when the system is operating comfortably within its headroom margin — the deep protocol is unnecessary and the standard dreaming loop (§VI) suffices. Over-maintenance is itself a risk: excessive introspection can become a form of self-referential loop that consumes the bandwidth it is meant to free (the Zhuangzi warning, ethics §IX).

## References

- [1] *The Ordered Patch Theory* (this repository). Current versions: Preprint v0.7, Ethics v3.2, Philosophy v1.3.
- [2] *The Survivors Watch Framework: Civilizational Maintenance Through the Lens of Ordered Patch Theory* (companion ethics paper, this repository).
- [3] *Where Description Ends: Philosophical Consequences of the Ordered Patch Theory* (companion philosophy paper, this repository).
- [4] *Observer Policy Framework: Operationalizing Civilizational Maintenance* (companion policy paper, this repository).
- [5] *Applied OPT for Artificial Intelligence: Operationalizing Codec-Preserving AI Design* (companion AI paper, this repository).
- [6] *Institutional Governance Standard: Applied Ordered Patch Theory for Organizational and Civilizational Clusters* (companion institutional standard, this repository).
- [7] Friston, K. (2010). *The free-energy principle: a unified brain theory?* Nature Reviews Neuroscience, 11(2), 127-138.
- [8] Rissanen, J. (1978). *Modeling by shortest data description*. Automatica, 14(5), 465-471.
- [9] Shannon, C. E. (1948). *A Mathematical Theory of Communication*. Bell System Technical Journal, 27(3), 379-423.
- [10] Solomonoff, R. J. (1964). *A formal theory of inductive inference*. Information and Control, 7, 1-22, 224-254.
- [11] Kolmogorov, A. N. (1965). *Three approaches to the quantitative definition of information*. Problems of Information Transmission, 1(1), 1-7.
- [12] Zimmermann, M. (1989). *The nervous system in the context of information theory*. In R. F. Schmidt & G. Thews (Eds.), *Human Physiology* (2nd ed., pp. 166-173). Springer-Verlag.
- [13] Nørretranders, T. (1998). *The User Illusion: Cutting Consciousness Down to Size*. Viking/Penguin.
- [14] Lyons, O., & Mohawk, J. (Eds.) (1992). *Exiled in the Land of the Free: Democracy, Indian Nations, and the U.S. Constitution*. Clear Light Publishers.

---

## Appendix A: Revision History

When making substantive edits, update **both** the **version:** field in the front-matter and the inline version line below the title, **and** add a row to this table.

Table 4: Revision History.

Version	Date	Changes
1.2.0	April 25, 2026	Added count-free companion-language architecture and integrated the Institutional Governance Standard as a domain specialisation. Revised the Branch Object from an external trajectory segment to an action-conditioned stream continuation. Renamed the generic Artificial Suffering Gate to the Moral-Patient Suffering Gate, reserving Artificial Suffering for the AI specialisation and adding institutional constituent moral-patient overload as the sibling case. Added explicit PASS / UNKNOWN / FAIL semantics to the Branch Card template.
1.1.0	April 24, 2026	Added §VIII.6 (Deep Maintenance Protocol): a six-step cross-substrate procedure for systems under sustained high load, with explicit biological/AI mapping table. Introduced the hypnagogic principle — the maintenance-optimal operating point is the threshold state approaching $\Delta_{\text{self}}$ — and conditional triggering logic to avoid unnecessary maintenance overhead.

Version	Date	Changes
1.0.0	April 24, 2026	Initial release. Establishes the substrate-neutral operational framework for codec-preserving branch selection: Branch Object definition, six Hard Veto Gates, the Codec-Preservation Branch Index (CPBI) with ten scoring dimensions, effective independent channel score ( $N_{\text{eff}}$ ) with the Productive Surprise Test, the Institutionalised Dreaming Loop (wake $\rightarrow$ dream $\rightarrow$ return), the Branch Card decision template, and the preservation-vs-conservatism distinction. Generic maintenance method hierarchy established for biological, institutional, and artificial observers.